Changes of AI Receptive Fields With Sound Density

DAVID T. BLAKE AND MICHAEL M. MERZENICH
Coleman Laboratory and Keck Center for Integrative Neuroscience, Department of Otolaryngology, University of California, San Francisco, California 94143 -0732

Received 29 March 2002; accepted in final form 14 August 2002

Blake, David T. and Michael M. Merzenich. Changes of A1 receptive fields with sound density. J Neurophysiol 88: 3409–3420, 2002; 10.1152/jn.00233.2002. Primates engage in auditory behaviors under a broad range of signal-to-noise conditions. In this study, optimal linear receptive fields were measured in alert primate primary auditory cortex (A1) in response to stimuli that vary in spectrotemporal density. As density increased, A1 excitatory receptive fields systematically changed. Receptive field sensitivity, expressed as the expected change in firing rate after a tone pip onset, decreased by an order of magnitude. Spectral selectivity more than doubled. Inhibitory subfields, which were rarely recorded at low sound densities, emerged at higher sound densities. The ratio of excitatory to inhibitory population strength changed from 14.4:1 to 1.4:1. At low sound densities, the sound associated with the evocation of an action potential from an A1 neuron was broad in spectrum and time. At high sound densities, a spike-evoking sound was more likely to be a spectral or temporal edge and was narrower in time and frequency range. Receptive fields were used to predict responses to a novel high-noise-density stimulus. The predictions were highly correlated with the actual responses to the 2-s complex sound excerpt. The structure of prediction failures revealed that neurons with prominent inhibitory fields had relatively poor linear predictions. Further, the finding that stochastic variance is limiting in prediction even after averaging 150 repetitions means that high-fidelity representations of simple sounds in A1 must be distributed over at least hundreds of neurons. Auditory context alters A1 responses across multiple parameter spaces; this presents a challenge for reconstructing neural codes.

INTRODUCTION

The auditory system successfully performs sound identification across signal-to-noise conditions ranging from quiet to complex noisy backgrounds and environments. To address what signal transformations contribute to such behavior, the changes in the auditory-system sound representation under such disparate noise conditions were measured using linear regression. Predictions were made and compared with actual responses to assess the validity of the linear model in characterizing neuronal function.

Different techniques for measuring the receptive field of neurons in primary auditory cortex of the awake primate yield conflicting results. If a high-spectrotemporal-density stimulus is used to estimate the receptive field through linear regression, neurons in A1 respond to a best 50-dB-SPL tone pip with changes in firing rate of up to 10 action potentials per second, but in general, with 1–3 imp/s (deCharms et al. 1998). Alternatively, if isolated tone pips are used, neurons in primate A1 respond with a response magnitude that is an order of magnitude greater (Recanzone et al. 2000). Both characterizations are valid within their contexts, but neither can stand alone as representative of A1 neuronal responses.

Here, the responsiveness and selectivity of neural responses in A1 were defined as the spectrotemporal sound density was systematically changed. All stimuli were composed of the same elements, tone pips, that were randomized in time across 84 1/12-octave bands. The average response to tone pips in each band was measured and corrected for stimulus correlations within and across frequency bands. The result was the optimal linear spectrotemporal receptive field as determined by linear regression (DiCarlo and Johnson 1999). The reverse correlation process also defines the optimal stimulus for a cortical neuron (deCharms et al. 1998), so changes in receptive field selectivity and responsiveness are mirrored by changes in the selectivity of the sound associated with action potentials.

After measuring the receptive fields under different sound-density conditions, the predictive power of the contextually dependent receptive fields was tested. Previous predictive work in the grass frog midbrain has demonstrated the poor power of the spectrotemporal receptive field (STRF) to predict responses to stimuli different from those used to derive the receptive field (Eggermont et al. 1983). In the avian forebrain, linear receptive fields optimized by regression have been shown to have more predictive power if the estimation stimulus is similar to the stimulus used for prediction (Theunissen et al. 2000). Even so, capturing half of the response variance, or a correlation coefficient \( r > 0.7 \) or \( r^2 > 0.5 \), is a challenge (Sen et al. 2001). Here receptive fields were defined at one sound density, and the responses of neurons to a novel repeating stimulus at the same density were recorded. System linearity was tested by comparing the response predicted from the receptive field with the actual response.

Our multichannel chronic implant recording preparation (deCharms et al. 1999) offers important advantages in providing access to a large number of neurons in the awake animal and in allowing adequate recording time with each unit to be able to deliver the required long stimulus sets, without limitations on recording stability or effects of anesthetic agents on the unit responses.

Systematic changes in A1 receptive fields are described that can explain the differences between reverse-correlation defined...
and isolated tone-pip-defined receptive fields. The prediction experiments provide insights into the sizes of the distributed responses necessary to reconstruct a stimulus within A1 and demonstrate that receptive fields with substantial inhibitory domains are less amenable to reverse correlation analysis.

METHODS

Physiological recordings

Data were obtained from three chronically implanted owl monkeys, *Aotus nancymae*. Implants were placed into the presumed primary auditory cortex (A1). The A1 target was 2–3 mm anterior interaural and just lateral to the temporal/frontal fissure in the lateral bank of the lateral sulcus. Transcranial recording through burr holes confirmed A1 response characteristics and expected tonotopy (Imig et al. 1977; Recanzone et al. 1999). Surgery was performed under areflexic barbiturate anesthesia. Techniques for implantation are described in a methods paper (deCharms et al. 1999). Histological confirmation of A1 was performed in one animal by identification of microelectrode tracks in areas of cortex with densely stained (cresyl violet) middle lamina.

Recordings were made with parylene-insulated iridium microelectrodes (Micro Probe, Potomoc, MD) with tip exposures between 5 and 7 μm long, chosen to maximize probability of sampling single units (Galambos and Davis 1943; Hubel 1957). Implant best frequencies spanned the range of frequencies found on the exposed surface ranging from 110 Hz to 20 kHz. Owl monkey A1 contains one higher unsampled octave, the representation of which lies entirely within the superior bank of the Sylvian fissure.

After implantation, a recovery period of several weeks ensued before recording was initiated. For recording sessions, the animal sat passively in a primate chair with its head reliably positioned approximately 24 inches in front of a free-field speaker. Animals were monitored in all recordings, and recordings with substantial head movements were discarded. No attentional control beyond the animal maintaining this head-positioning standard was used. Single units were isolated on-line using the Magnet system (Biographics, Winston-Salem, NC), and 1.5 ms of spike waveform were stored for each unit discharge event, beginning 1 ms before a voltage threshold crossing. Off-line, single-unit quality was confirmed by a waveform analysis that used three criteria: signal-to-noise ratio, coefficient of variation (CV) of maximal positive slope on the principal waveform deflection, and CV of maximal negative slope on the principal deflection. The signal-to-noise ratio, or the mean peak-to-peak magnitude divided by the noise SD, had to exceed 5, and CVs for each unit had to be below 0.25. Multunit recordings consisted of recordings that were manually selected as single units that did not meet our single-unit statistical criteria. Receptive field spectrotemporal density analysis used only single-unit recordings. To increase statistical power, prediction experiments used both single- and multunit recordings.

Sound presentation

Sound levels were calibrated with a Bruel and Kjær sound level meter using the “A” filter. Sounds were created digitally, recorded on an audio CD, and played through a McIntosh audio amplifier. Each sound element was a 20-ms-duration tone pip that was 50 dB SPL in amplitude, with 5-ms raised cosine onset and offset ramps. The onset ramps can be described by the equation \((1 - \cos(2\pi t/10 ms))/2.0 \) for \(0 < t < 5\). Offset ramps are the time reverse of onset ramps.

For the purposes of this study, the basic tone pip was considered an approximation of a frequency-specific impulse function. A linear system’s response to an impulse function is complete in its temporal characterization of the system. Although one impulse response would be enough to characterize a noiseless system, the primate auditory cortex is stochastic in its response.

Multiple impulses were presented in each frequency band to improve estimates of the temporal response function within each frequency band. For all stimuli, every 1/12th octave frequency band contained an independent Poisson train of tone pips with the same mean rate of tone pip presentation. A Poisson train of tone pips was used because it samples all tone pip repetition frequencies equally.

Each spectrotemporal density stimulus was created by adding the independent tone pip trains at each 1/12th octave. An octave is a doubling of frequency. One second of each stimulus is shown in Fig. 1. The spectrotemporal densities used were one tone pip per octave per 720, 225, 64, 20, and 8 ms. Previous reverse correlation studies by our group used the second most dense sound (deCharms et al. 1998). Each stimulus was presented continuously for 5 min, with 1 s of silence between stimuli. Single neurons were sampled with the first stimulus set for 25 min.

For the second stimulus set, only one spectrotemporal sound density was used: one tone pip per octave per 64 ms or ordinal density 3. The stimulus consisted of 5 min of the random tone pip stimulus, followed by an independent 2-s segment at the same spectrotemporal density repeated 150 times. A subset of neurons was sampled using both stimulus sets. Microelectrodes in the implant were not moved for months; sampling single neurons for several hours was routine.

A third stimulus set was a standard tuning curve stimulus set. The 84 frequencies were presented at eight intensities from 10 to 80 dB SPL as 50-ms tone pips with 5-ms raised sinusoid ramps. Stimulus onsets were separated by 200 ms. Each intensity-frequency combina-

![Fig. 1. Spectrograms of stimuli at 5 spectrotemporal densities. Each horizontal black bar represents a 20-ms tone pip at its time and frequency position. Within a single plot, the temporal density of tone pips at each frequency is the same. In the 5 plots, for each frequency, the average tone pip rate was 1 pip per 8,640, 2,700, 768, 240, and 96 ms. These densities can also be described as tone pips ⋅ octave⁻¹ ⋅ time⁻¹ as described in METHODS.](image-url)
tion was presented four times in randomized blocks, so that the entire stimulus set lasts about 9 min.

**Receptive field estimation**

A technical description of the methods for this section has been described for receptive field estimation in the somatosensory system (DiCarlo and Johnson 1999). The technique measures the linear filter optimized by linear regression that matches the response properties of the neuron. This filter is also the least-squares estimate of the linear receptive field (Jackson 1989). In practice, this technique may be used for any stimulus for which the stimulus autocorrelation matrix is full rank, i.e., every stimulus element used in the regression is presented independently from all other elements.

For each stimulus, the response was assumed to come from a basic linear model

\[
\text{[Stimulus]}[\text{RF}] = \text{[Response]}
\]

where the stimulus is a 60,000 × 1,680 array. There are 20 columns for each frequency that cover 50-ms increments in latency up to 100 ms. Each 5-ms bin contains the number of tone pip onsets that occurred in that period. The receptive field is a column vector that contains an element for each frequency and each 5 ms up to 100 ms, or 1,680 elements in a single column. The appropriate solution for the receptive field is

\[
\text{[RF]} = \text{[([Stimulus])} \cdot \text{[Stimulus]} \cdot \text{[Stimulus]} \cdot \text{[Response]}
\]

As each of the 84 stimulus bands was allowed to interact with the receptive field for 100 ms, or 20 time bins, the stimulus autocorrela-

tion matrix was 1,680 rows and columns. Matlab (Mathworks) was used to perform the matrix inversion. All other analysis was done using the C programming language and the GNU C compiler. The receptive field estimation is equivalent to performing a linear regression on the 1,680 unknowns. The Response is 60,000 rows and 1 column of spike occurrences in 5-ms time bins.

The [Stimulus] [Response] contains the cross-products of the 1,680 columns of the stimulus matrix with the rows of the response vectors or the scaled temporal responses to each tone pip. These cross-products were calculated first with 1-ms resolution. This calculation was equivalent to counting the action potentials in each millisecond after the occurrence of tone pips in each frequency band. This raw product was smoothed using a Parzen windowing algorithm. In this algorithm, each spike is replaced by a Gaussian with its SD inversely proportional to the spike rate. Spike rates more than 5 spikes/s were assigned a Gaussian with a SD of \(\frac{20}{\text{rate}}\); spikes in regions with lower spike rates were assigned a Gaussian with a 5-ms SD. This method allowed a tradeoff between uncertainty in spike frequency and spike timing based on the Gabor uncertainty principle (Gabor 1946). Parzen binning of two-dimensional firing rate arrays has been used to mini-

mize SEs in estimating the firing rates (Blake et al. 1997b). The 1-ms bins were then compressed into 5-ms bins to form the final reverse correlation product [Stimulus] [Response].

The stimulus autocorrelation matrix was then inverted. In all cases, the stimulus autocorrelation matrix was positive definite and strongly diagonalized (Horn and Johnson 1985) so that there a robust inverse matrix was attained. The principle difference in the autocorrelation matrix across stimuli was that nondiagonal terms increased relative to diagonal terms as the stimulus became more dense, although in all cases, diagonal terms were much larger. The nondiagonal terms are the probabilities of tone A occurring within 100 ms of tone B, if A and B are different tones. As each stimulus was an independent Poisson train of tone pips, little structure existed in the nondiagonal elements of the autocorrelation matrix. The next step was left-multiplying the autocorrelation inverse with the [Stimulus] [Response] to achieve the result of the linear filter optimized by linear regression to match the function of the neuron. This computation is also the linear filter achieved by minimizing the squared error between the response predicted by the linear filter and the actual response. Note that each receptive field has an offset term that corresponds to the average firing rate divided by the mean tone pip presentation rate. In each receptive field, a row corresponded to the expected response of a neuron to a single presentation of a tone pip in the context of the density in which the receptive field was measured.

**Extraction of measures from the receptive field**

To extract the measures from the receptive field, the first determina-
tion was whether or not a receptive field was present. The second step was defining what should be included in the receptive field. Completely automated quantitative criteria for these steps were used to avoid potentially subjective biases.

**Determination of whether a receptive field was present**

One method that can be used for determining if a receptive field is present is comparing the variance in the firing rates to the variance in smoothed receptive field rates (DiCarlo et al. 1998). Response estimation errors in adjacent bins in reverse correlation are largely un-
correlated, if the errors occur from stochastic variation. Smoothing the receptive field removes more reverse correlation errors than it does the receptive field. The smoothing function must be narrow relative to expected receptive field structure. If the smoothing removed too much of the firing rate relative to the maximal rate, the receptive field was removed from further analysis. In our sample, receptive fields were smoothed with a Gaussian whose SD was half the bin size. The root mean squared difference between the smoothed and unsmoothed receptive fields was the removed noise measure. Also, the maximal unsmoothed firing rate was determined over all frequencies tested from time 0 to 100 ms after tone pip onset. If the ratio, was less than 12% for excitatory components, then the receptive field was assumed to included an excitatory component. The ratio was larger for inhibitory components, 18%, due to a floor effect on the negative values. The floor effect occurred because negative receptive field values may only be observed up to values close to the average firing rate. This portion of the analysis was carefully constructed because receptive fields change with sound density, which causes nearly all neurons to assume receptive fields close to the limits of the noise in the regression process. Threshold values, 12 and 18%, were selected by manual comparison of receptive field images and ratios over a substantial portion of the sample. The threshold was set to be conservative and avoid false positives but otherwise include as many receptive fields as possible.

The removed-noise measure had a strong tendency to decrease with increases in spectrotemporal density. The mean number of presenta-
tions of tone pips per frequency band were 35, 111, 391, 1,250, and 3,125. The number of action potentials contributing to a receptive field may be estimated as the product of the strength measure and the number of tone pip presentations i.e.,: at the lowest tone pip density, approximately 120 action potentials (1 per 2.5 s) contributed to the receptive field, and at the highest density, approximately 1,300 action potentials (4/s) contributed. These numbers are averages; actual per unit data on receptive field strength are presented in results. This number indicates the number of action potentials that occurred in a time-structured relationship with tone pips that allowed the estimation of an expected response.

The number of PSTHs averaged to form a single row of the receptive field changed by a factor of 90. The noise should decrease roughly with \(\sqrt{\text{number of presentations}}\), or by a factor of 9.5, if the receptive fields are unchanged by stimulus density. If, instead, the receptive fields change with density, the relative noise change should be a product of the \(\sqrt{\text{number of presentations}}\) and the ratio of the process SD and its mean. As the measure is an action potential
counting process, the ratio between its SD and mean tends to increase as the counts decrease. A rough estimate would be that the mean decreases by a factor of 10, and the ratio of the SD to the mean increases by a factor of 3. So, the change in trial number decreases the noise by 9.5, the change in neural sensitivity decreases the signal by 10 and increases the relative noise by a factor of 3.

The changes in noise raise the issue of a bias in selection of receptive fields at different densities. The use of more tone pip presentations at higher densities decreases the estimation standard errors. If a receptive field is constant across densities, detecting it will be easier at higher densities. If the ratio between response sensitivity and receptive field noise decreases, the receptive field is less detectable. Excitatory fields were actually found to be less detectable at higher densities. This effect is caused by the relative noise increase. Inhibitory fields were more detectable at higher densities, which would be predicted by the change in noise with density alone, although there were also decreases in inhibitory rates with increasing density.

**Determining a threshold for including a bin in the receptive field**

A widely used technique for thresholding a receptive field is to determine a maximal rate measure, then to threshold the receptive field at a fraction of that maximal rate (Blake et al. 1997a,b; DiCarlo et al. 1998; Johnson and Lamb 1981; Phillips et al. 1992). This technique creates a receptive field size estimate that is unbiased by the size of sampling. In our case, the largest $2 \times 3$ bin average was used as the maximal rate measure; the threshold was 50% of this measure. In addition, to rid the receptive field of singular noise bins, any part of the receptive field had to have two of the eight surrounding bins included in the receptive field or be at least 85% of the maximal measure. The fraction of the maximal measure used for thresholding was derived independently for excitation and inhibition because inhibitory bins again suffered from a floor effect. The 50% criterion was derived by comparing unthresholded and thresholded receptive fields over most of the sample to minimize false positives. A thresholding example is shown in Fig. 2.

Receptive field areas are proportional to the number of time-frequency bins in the receptive field. Excitatory strength is the sum of all positive time-frequency bins in the receptive field, and the sum of negative time-frequency bins is inhibitory strength.

These methods for determining and thresholding a receptive field were tested. Simulated linear neurons with simple excitatory fields were convolved with the stimuli to generate responses. Such linear neurons had identical receptive fields at different spectrotemporal densities.

**Prediction**

For the prediction, receptive fields were estimated using the first 5 min of responses to the third spectrotemporal density. The next 5 min was 150 repeats of a 2-s excerpt, and the average response to the 150 trials was compared with the response expected from the receptive field convolution with the stimulus. Receptive field estimation was done as described in the preceding text, except that no Parzen windowing procedures were used. The 2-s peristimulus time histogram (PSTH) was binned in 5-ms bins. In this procedure, action potentials occurring in each bin were added together to form the total for that bin.

The prediction was created by calculating the receptive field and circularly convolving it with the 2-s repeating stimulus. In this procedure, for each tone pip, the row of the receptive field corresponding to that frequency was added to the PSTH shifted by the time of the tone pip. Negative values were outside the range of the model and were clipped to zero after the convolution. The sum of the contributions of each tone pip was then evaluated against the actual response. The use of the term percent explained variance in the text indicates the $r^2$ linear correlation coefficient. This is the percent of variance in the actual response firing rates that can be accounted for by correlation with the predicted response.

**Analysis of prediction failures**

The hypothesis that the history of firing of the neuron impacted deviations from linearity in the prediction was also tested. To consider this, assume the response is of the form

$$R(t) = LR(t) + f(R(t - 1), R(t - 2), \ldots, R(t - 10))$$

and

$$f(R(t - 1), R(t - 2), \ldots, R(t - 10)) = a_{11}R(t - 1) + a_{21}R(t - 2) + \ldots + a_{101}R(t - 10)$$

In this formulation, $R$ is the actual response, $LR$ is the linear prediction, and $f()$ is the function of the past history of the neuron. The coefficients for $a_{11}, \ldots, a_{101}$ can be determined using linear regression techniques similar to those described for calculating the receptive field, and $R(t) - LR(t)$ is the residual of the estimation.

Last, a step-wise regression was used to determine variables that correlated with the goodness of fit of the prediction. Details are included in the text.

**RESULTS**

Two exemplary sets of A1 single-unit receptive fields are shown in Fig. 3. The stimuli for these receptive fields, shown in Fig. 1, are sums of randomly distributed tone pips. The representative neuron in Fig. 3A responded most strongly to a tone close to 440 Hz. As sound density increased, the response to the best-frequency tone pip decreased from 150 to about 5 spikes per tone pip. The number of spikes generated in response to a best tone pip decreased from about three spikes per tone at the lowest density to one spike per 10 tones at the fourth density. This example neuron did not have a measurable receptive field at the highest sound density, using the statistical criteria described in METHODS.

The form of this neuronal receptive field and the total area of the excitatory and inhibitory fields changed with sound density. Spectral selectivity was reduced from more than 1/2 to about 1/4 octave. Spectral selectivity may be observed in this plot as the vertical extent of the red excitatory component of the receptive field. It is less clear in this example if the temporal selectivity for the excitatory component changed as well as was often clearly the case.

Another change in form of this representative example was
the emergence of inhibition at the third and fourth spectrotemporal densities. As the density increased between those two stimuli, the receptive field inhibition halved in strength, while the excitatory component decreased by a factor of four in intensity, and decreased in spectral extent. This example was dominated by excitatory components at the lowest sound densities and was more closely balanced with excitatory and inhibitory components at the higher sound densities. Note that we refer to components associated with a decrease in firing rate as “inhibitory” components and components associated with increases in firing rate as “excitatory” components. These terms are derived from measurements of action potential counts; they do not directly imply synaptic sources.

Figure 3B shows a second representative example of receptive field variation with sound density. This example also changed in size and peak intensity. However, this neuron, as for approximately half of the neurons in our sample, did not develop a linear inhibitory component at higher sound densities.

As shown in Fig. 4, the effects of sound density on the maximum response rate in the receptive field were fairly homogeneous. An ANOVA on the effects of sound density on peak rate revealed that these changes were statistically signifi-
The number of sound densities that yielded a measurable receptive field varied substantially from neuron to neuron. Of all 191 neurons that had measurable receptive fields at one or more densities, 90 had inhibitory receptive fields at one or more densities. All neurons had measurable excitatory components at one or more sound densities, although the selection criteria included any neuron with a receptive field irrespective of its sign. Table 1 shows the percentage and number of recorded A1 neurons that had measurable receptive fields at different sound densities. There was a statistically significant trend for inhibition to favor higher sound densities, and excitatory receptive fields to decrease in probability as the sound density increased. The statistics in this case are similar to the case of a loaded coin flip in which “heads” corresponds to a structured receptive field being present, and “tails” corresponds to its absence. The largest standard error for such a Bernoulli model of receptive field presence was 3.6% for the fourth excitatory sound density. Using Bernoulli distribution functions, all changes in percent of neurons having structured excitatory receptive fields at different spectrotemporal densities were significant at the 5% level. For inhibitory fields, all changes in density except the lowest two had statistically significant differences.

The single neuron area data for all nonzero receptive fields are shown in Fig. 5A. Each red line shows adjacent nonzero receptive field excitatory areas, and each blue line shows nonzero inhibitory areas. The average and median area of all nonzero receptive fields is shown in Fig. 5B. Inhibitory components were large relative to excitatory components and decreased in size with increases in spectrotemporal density above the third density.

All areas for each polarity and spectrotemporal density were not significantly different from log normal distributions using one-tailed Kolmogorov-Smirnov tests. All excitatory area distributions, and the fourth and fifth density inhibitory area distributions, had significant skewness relative to the standard normal distribution (P < 0.001). The structure of the distribution is of interest because it allows inferences as to how the distribution is created. A plot of the area distributions for each polarity and spectrotemporal density is shown in Fig. 5C. Each red line plots the distribution of excitatory areas for one spectrotemporal density. The three blue lines plot the third, fourth, and fifth spectrotemporal density inhibitory area distributions. Means and statistics were not computed on the lowest two densities for inhibitory data because there were only five neurons that had inhibitory fields at the low densities. For the plot, each distribution was log transformed. Its mean was then subtracted, and its values divided by its SD. If the data were log normal, the remaining distribution will be standard normal. The black line plots the standard normal distribution. The log normal s shape parameters, or the multiplicative SD (Limpert et al. 2001), ranged from 1.51 to 1.80 for the excitatory regions and 1.23 to 1.42 for the inhibitory regions. An s shape parameter of 1.0 corresponds to a standard normal distribution; larger s parameters correspond to longer tailed distributions.

The significance of the changes in area with density is summarized in Table 2, which shows t-tests significant with P < 0.01 for the log transformed nonzero areas. There was a trend for the lowest-sound-density areas to be larger than the

<table>
<thead>
<tr>
<th>Sound Density</th>
<th>Percent Excitatory</th>
<th>Number Neurons</th>
<th>Percent Inhibitory</th>
<th>No. of Neurons</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (low)</td>
<td>95.3</td>
<td>182</td>
<td>2.6</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>85.3</td>
<td>163</td>
<td>1.6</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>73.8</td>
<td>141</td>
<td>11.5</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>52.8</td>
<td>101</td>
<td>34.0</td>
<td>65</td>
</tr>
<tr>
<td>5 (high)</td>
<td>38.7</td>
<td>74</td>
<td>22.5</td>
<td>43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sound Density</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FIG. 5. Effects of spectrotemporal density on receptive field area. One hundred ninety one neurons yielded a measurable receptive field at 1 or more sound density. A: single neuron data. All single neuronal data points containing excitatory receptive fields at adjacent spectrotemporal densities are represented by red lines; adjacent inhibitory area data points are represented by blue lines. B: the average and median of all nonzero receptive fields is shown for each density. The ordinate for the inhibitory components has 3 times the scale of the excitatory components. Data points are omitted for the 2 lowest densities of inhibition because the samples were too small. Error bars show the SE. C: normalized area distributions compared with the log normal distribution. For each spectrotemporal density and polarity (excitatory or inhibitory), the areas were log transformed, mean subtracted, and normalized by the SD to compare against the standard normal distribution shown with the black line. None of the distributions was significantly different from the log normal.
higher-sound-density areas. The tables are diagonally symmetric, and only the upper right half is used. For inhibitory area data, statistics were computed comparing the third, fourth, and fifth densities. The inhibitory area at the third density was significantly smaller than those at the fourth \(t\text{-test, } P < 0.01\).

Although these measures test for changes in all sampled structured receptive fields at each density, they do not address the issue of whether single neurons also change, because they do not control for changes in sampled sets of neurons. To address that issue, the neuronal set was subsampled for neurons with excitatory receptive fields at each pairwise comparison. For example, to compare density one and density two, the samples are subsampled for neurons with areas at both densities, and then a \(t\text{-test} is performed. Pairwise \(t\text{-tests} found significant decreases in area with density \(1 > 3\ (P < 0.01\). Paired neuronal samples were similarly subsampled for comparisons of inhibitory areas, but no significant effects were found. It should be noted that such subselection reduced the sample size and also the statistical power.

To determine changes in temporal and spectral selectivity, receptive field areas were analyzed for temporal and spectral range. The 10–90% range was used as the measure of bandwidth. Temporal selectivity is shown in Fig. 6, A and B. In Fig. 6A, all adjacent nonzero data points are plotted for each single neuron. The mean excitatory temporal bandwidth decreased from 40 ms at the lowest sound density to 23 ms at the fourth density. The lack of a direct correspondence between the area plots in Figs. 5 and 6 is due to the irregular shapes of receptive fields, and long-tailed distributions of these parameters. Both spectral and temporal bandwidth means were substantially larger than the medians shown with the dashed lines in Fig. 6B. Table 3 shows the results of \(t\text{-tests} significant for \(P < 0.01\) comparing the nonzero means of each group. No significant differences were found between the third, fourth, and fifth inhibitory density temporal bandwidths.

<table>
<thead>
<tr>
<th>Sound Density</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If the temporal bandwidth data are subsampled for single neurons that had excitatory fields at the compared densities, there are significant differences between the first and third density only \(t\text{-test, } P < 0.01\). No significant differences are found if the inhibitory data are subsampled.

The plots for spectral bandwidth are shown in Fig. 6C and D. The mean excitatory bandwidth changed from greater than 2 octaves at the lowest sound density to less than 0.8 octaves at the fourth sound density. The excitatory medians ranged from 1.2 to 0.42 octaves, and the inhibitory medians were 1.5–2.0 octaves. Table 4 shows the results of \(t\text{-tests} significant for \(P < 0.01\) comparing the nonzero means of each group for the excitatory and inhibitory components. Again, the lowest spectrotemporal density bandwidths tended to be broader in selectivity than the higher sound densities. Inhibitory spectral bandwidths at the third, fourth, and fifth densities were not significantly different. If these excitatory spectral bandwidths are subsampled for neurons having excitatory fields at the three lowest densities, the pairs \(1 > 2\) and \(1 > 3\) are again significant \(t\text{-test, } P < 0.01\). The subsampled neurons with inhibitory fields were not significantly different.

The total strength and balance of excitatory and inhibitory receptive fields was also considered. For this analysis, each neuron was considered as a member of the total pool, and averages were derived over all neurons that yielded measurable receptive fields for any sound density. The plot of the strength
against the ordinal spectrotemporal density is shown in Fig. 7. Although excitatory contributions decreased monotonically with increases in spectrotemporal density, inhibitory strength did not change. Pairwise $t$-tests confirmed all changes in sound density, except differences between densities 4 and 5, had significant effects for $P < 0.0001$ on the mean excitatory strength, whereas no changes in sound density had a significant effect on mean inhibitory strength. Further, the balance of excitation and inhibition changed with sound density from a ratio of 14.4:1 at sound density one, to 1.4:1 to 1.5:1 at the two highest sound densities. Excitatory and inhibitory strengths were also not significantly different from log normal distributions. All distributions, except the third density inhibitory strength, had significant skewness relative to a standard normal distribution ($P < 0.001$). If the total strength data are subsampled for neurons having excitatory fields at combinations of two densities, lower density excitatory strengths are significantly smaller ($t$-tests, $P < 0.001$) except the comparison between the fourth and fifth density. If the inhibitory strength is subsampled for neurons having inhibitory densities at the third and fourth densities, all neurons have inhibitory strengths greater at the third density than at the fourth density, and this change is significant (sign test, $n = 15$, $P < 0.0001$). The inhibitory population totals are not significantly different because the decrease in single neuronal strength was balanced by the increase in proportion of neurons having structured inhibitory receptive fields.

Prediction of responses to novel stimuli with context dependent receptive fields

We have performed comparisons on the spike count and selectivity of responses to 50-dB tone pips presented either as part of the lowest density stimulus set or as part of a tuning curve stimulus set. The responses, while being highly stochastic, were all within the appropriate range of selectivity and response rate to have been generated by the same response functions. The tuning curve repetition rate was one tone pip per octave per 1,400 ms compared with one tone pip per octave per 700 ms for the lowest-density random tone pip stimulus.

To determine if the responses to the denser stimuli were also representative of the neuronal responses at those densities, a second stimulus set was used. The first 5 min of the sound stimulus set contained randomized tone pips at the tone pip density of one pip per octave per 64 ms, i.e., ordinal density 3. The second 5 min consisted of a 2-s sound repeated 150 times. That 2-s segment consisted of randomized tone pips delivered at the same spectrotemporal density as the first 5 min. Receptive fields were measured during the first five minutes. A predicted response to the 2-s segment was created by convolving the receptive field over the sound. Neurons with sustained rates to the 2-s segment of less than 5 spikes/s were omitted. In addition, the true response to the first 75 trials was computed, and correlated with the true response to the second 75 trials, and any neurons with $r^2$ values less than 0.50 were omitted. These criteria were established so that only neurons with large and adequately consistent responses were evaluated. Neurons with $r^2$ values less than 0.5 had receptive fields with no obvious predictive power. After applying those criteria, single- and multunit response were combined for greater statistical power. There were no obvious differences between the single- and multunit groups in prediction factors or magnitudes.

The example in Fig. 8 shows two single units with receptive fields high consistency. The example on the left additionally had one of the best fits of any neuron to the actual recorded response; the neuron on the right was the worst fit of any neuron with a high consistency. The remainder of the analysis consisted of attempts to explain why some receptive fields predicted better than others.

In looking for the source of residual lack of fit, the first attempt was to look for firing rate history effects. The residual is the true PSTH firing rates minus the predicted PSTH firing rates. A model was generated as described in METHODS. The model tested the assumption that some factor added to the current firing rate was based on a linear combination of previous rates. Although the fits were significant, only the first bin in the firing rate history had a significant effect, and it was positive. This means that a high firing rate 5 ms in the past predicted a high residual now. However, the residual contained much of the variability of the true PSTH, so that some of this predictability was an artifact of the lack of fit of the linear prediction. Further, the tone pips were not aligned at 5-ms intervals, so the predictive reconstruction would smear firing rates from 5-ms bin to bin.

Another strategy to explain variance in neuronal fits was to perform a step-wise regression to relate variables associated with the neurons to the goodness of fit of the model. In step-wise regression, all tested variables are correlated with the

![FIG. 7. Effects of sound density on response strength. One hundred ninety one neurons yielded a measurable receptive field at 1 or more sound density. A: single neuron data. All adjacent nonzero values for single neurons are shown connected with red lines for excitatory strength, and blue lines for inhibitory strength. B: the average of all receptive fields, including 0 size fields, is shown for each density. Error bars show the SE.](http://jn.physiology.org/Downloaded.from.http://jn.physiology.org/ by 10.220.33.5 on May 28, 2017)
The best tested variable is considered the largest source of variance. Further variables are then added to the regression in linear combination with the first factor. At each step, the best variable added will be kept, provided the improvement in explained variance is significant. The procedure is stepwise until no further variables significantly correlate with the residual, and removal of any variable results in a significant reduction in fit. The tested variables included excitatory and inhibitory strength, consistency, and mean rate.

The largest source of variance was the excitatory strength. This variable explained 55.5% of the variability in goodness of fit. The second largest source of variance was the neuronal consistency, which explained another 9.8% of the variance. This measure is the correlation of the first 75 trials response with the response to the next 75 trials. As first factors, consistency and excitatory strength were almost equal in contribution. Either variable explained more than 53% of the variability in fit when taken as a first factor.

The third largest source of variance was inhibitory strength. The inhibitory strength explained another 8.3% of the variance; its correlation with the goodness of fit was negative. Neurons with a lot of inhibition had poor fits, after adjusting for the response consistency and excitatory strength. One example of a neuron with high consistency and poor fit is shown in Fig. 8B.

The relation of each factor is shown compared with the model goodness of fit in Fig. 9A. Mean response rate was not a significant source of variability after adjusting for the first factors.

In Fig. 9B, the response rate predictions are repeated using the receptive fields from the lowest spectrotemporal density to predict the responses to the repeating novel stimulus that used the third spectrotemporal density. The first spectrotemporal density is close to the densities typically used in tuning curve stimuli; it evokes firing rates in the same range as conventional tuning curve stimuli. The receptive field measured at the third density had more predictive power than predictions at the first density in 40 of 43 cases ($P < 0.0001$, sign test). The correlation measure is insensitive to scaled changes, so the decrease in correlation exists without consideration of significant scaled differences in firing rate at the different densities.

To summarize, the average correlation of the first 75 trials with the second 75 trials was $r^2 = 0.7387$, although this only includes neurons with $r^2$ values more than 0.5. The average correlation between model and actual response was $r^2 = 0.3833$. The lack of fit between model and actual response was
explainable by three factors: poor neuronal consistency, excitatory strength, and inhibitory strength.

**DISCUSSION**

Our main finding is that there are systematic changes in receptive field structure as a function of the stimulus environment. Sounds associated with action potentials from cortical neurons change on a large scale as a function of stimulus context. The auditory system responds to a single tone pip with increased specificity, and by fewer evoked action potentials, as the sound density increases. We took advantage of this systematic change to demonstrate a nonlinear systems analysis technique, i.e., to determine analytically a steady-state operating point, then find the linear system closest to that operating point. The operating point, or tone pip density in our case, was used to classify the system, or receptive field, and to predict responses to stimuli close to that operating point.

Other studies have assessed the linearity of A1 cortical neurons. A shift in rate level functions with the addition of background white noise has been described (Phillips 1990). Nonlinear responses to combinations of two tones (Brosch and Schreiner 2000; Brosch et al. 1999) and nonlinear responses to combinations of up to nine tones (Nelken et al. 1994a,b) have been described. In these studies, the facilitation or depression of responses to a best tone have been elicited by tones outside the time and frequency response field of the neuron. The supralinear summation properties of cat A1 neurons in these studies have been compared with sublinear summation in the medial geniculate nucleus, inferior colliculus, and cochlear nucleus (Watanabe and Katsuki 1974). One study (Kowalski et al. 1996) has demonstrated that responses to linear combinations of sound stimuli, specifically moving ripple stimuli, are predicted by linear combinations of the responses to the ripple components. Nonlinear combination sensitivity has also been described for a variety of neuroethologically relevant stimuli in the nonprimary auditory cortex (Edamatsu et al. 1989; Suga et al. 1983; Taniguchi et al. 1986; Tsuzuki and Suga 1988) and is apparently contributed to by facilitatory influences through combination sensitivity in primary auditory cortex (Fitzpatrick et al. 1993; Kanwal et al. 1999; Misawa and Suga 2001). The ultimate goal of this work is to predict responses of A1 neurons to arbitrary stimuli. That goal will require many further studies that incorporate other stimulus attributes such as binaural stimulation (Kelly and Judge 1994; Miller et al. 2001; Semple and Kitzes 1993), different sound levels (Phillips 1990), and continuous tones (Ramachandran et al. 1999; Recanzone et al. 2000). These studies may be facilitated by the use of different basis sets (Calhoun and Schreiner 1998; Kowalski et al. 1996; Miller et al. 2002; Schreiner and Calhoun 1994; Versnel and Shamma 1998). In all cases, a close examination of nonlinearities can be facilitated with a predictive stimulus set to test the validity of modeling the neuron by analysis of its response properties.

The changes in receptive field strength and structure with changes in sound density are substantial, and several mechanisms may contribute. The hypothesis we favor is that the receptive field is composed of the sum of many inputs, each of which travels through synaptically coupled pathways from the cochlea to A1. If the synaptic depression is variable across the inputs that compose the receptive field, increases in sound density will change the relative contributions of those inputs, and the receptive field will shrink. An alternative viewpoint is that as the total stimulus intensity increases, the cochlea performs rate compression and shifts its thresholds for responses to individual tones upward (Gibson et al. 1985). The auditory system performs as though it was receiving input at a lower sound level (Phillips 1990), and receptive fields shrink because cochlear afferents have smaller receptive fields at lower sound levels (Galambos and Davis 1943). By that hypothesis, the rise in inhibitory strength is still unexplained.

The increases in the proportion of neurons that have measurable inhibitory receptive fields at high sound densities is probably a reflection of both the increased average firing rates at higher sound densities and less depression at inhibitory synapses than at excitatory synapses. If spontaneous rates are not substantial, inhibition may only be revealed by combining sound stimuli. Using a dense spectrotemporal stimulus to reveal inhibition is similar to previous efforts in the auditory system that present a best stimulus in conjunction with all other stimuli to assess an inhibitory field (Galambos and Davis 1944; Sachs 1969; Sachs and Kiang 1968). Recent studies (Galarreta and Hestrin 1998; Varela et al. 1999) have concluded that inhibitory synapses depress less than excitatory synapses, which would contribute to the changed balance of excitation and inhibition in receptive fields as sound density increases.

**Cortical representations and perceptual constancy**

This work has focused on the relationship between the acoustic stimulus and the receptive field. It may also be interpreted to imply the basis functions by which the animal hears sounds. Each action potential from each neuron told the animal, in a probabilistic sense, what sort of sound had just occurred. This message, as inferred from the receptive field, changes with the sound density. In a high-noise-density environment, the sound associated with a neuronal response is more spectrally and temporally restricted and can represent edges in time or frequency if the receptive field contains juxtaposed excitatory and inhibitory components. In a quiet environment, the sound associated with a response of the same neuron is less restricted spectrally and temporally than edges in time or frequency. Many behaviorally relevant audio signals are narrowband, such as many speech elements, marmoset twitter calls (Wang et al. 1995), and Mozart’s “Sonata for Two Pianos” (Rauscher and Shaw 1998). This differential selectivity in high-sound-density environments would have advantages in extracting narrowband signals from noise. Further, the sound associated with an action potential is much stronger in a high spectrotemporal density environment than in a low one. It remains to be seen how the animal can form an internal representation of the sound environment when the message from single neurons changes depending on the background noise context.

**Use of linear reconstructions to study cortical responses**

As linear receptive field estimation through reverse correlation becomes common in sensory cortices (DeAngelis et al. 1993; deCharms et al. 1998; DiCarlo et al. 1998; Jenison et al. 2001; Jones and Palmer 1987; Kowalski et al. 1996; Miller et al. 2001, 2002; Reid et al. 1997), the field needs to remain
cognizant that the method must approach a regressive estimate of a true receptive field to maximize predictive power. In all such measures, a consideration of the contributions to estimation errors can help interpret the results. Estimation errors dominate the structure of the smaller values in the optimal linear receptive field; the estimation power may be than signal to the prediction.

Neurons with inhibitory components had receptive fields that generated worse predictions than neurons without inhibitory fields. The measurement of inhibition is limited by the excitatory drive of the neuron because neurons cannot assume negative firing rates. For that reason, investigators have presented a strongly excitatory stimulus in combination with other stimuli to measure inhibition (Galambos and Davis 1944; Sachs and Kiang 1968; Sachs 1969). The poor predictive power of our receptive fields with inhibition is probably a reflection of the variable excitatory drive delivered to the neuron while inhibition is defined. This inadequacy would become important in reverse correlation schemes in which responses to physically different stimuli, e.g., light and dark spots, are mathematically considered to have opposite effects on cortical neurons (DeAngelis et al. 1993; Reid et al. 1997). Trying to match actual neuronal responses with predictions generated from receptive fields is an easy check on analytic techniques.

We also found that neurons with low consistency measures for the 2-s stimulus repeated 150 times predicted less than 20% of the response variance. Even among neurons for which the consistency measure exceeded 0.5, response consistency was strongly related to the goodness of the prediction. That a 150-trial estimate is limiting in the ability to predict responses indicates that reconstruction of the stimulus from the response on one trial must require a collective representation of inputs by hundreds to thousands of neurons.

Log normal distribution

The finding of a log normal distribution of receptive field area and strength comes on the heels of another study with similar findings conducted in the primary somatosensory cortex of the awake macaque (DiCarlo et al. 1998). Simple cell receptive fields in anesthetized cat V1 also appear to be long-tailed (DeAngelis et al. 1993). In general, a normal, or Gaussian, distribution is the result if some measure is the sum of many independent random events, none of which dominate the overall variance. A log normal distribution is the result if those small independent random events are multiplicative. Such distributions have been found for neuronal thresholds in auditory (Katsuki et al. 1962) and mechanoreceptive first-order afferents (Johnson 1974). An interpretation in these systems is that there are successive stages of attenuation that have independent sources of variability. In the case of CNS receptive field areas, however, it is less clear what the underlying multiplicative events are. One possibility is that the magnitude of receptive field change due to development or adult plasticity is proportional to the receptive field size. As receptive field strength also has a log normal distribution, we hypothesize that every action potential contributes equally to receptive field change, and that this mechanism leads to a log normal distribution of receptive field areas and strength.

Conclusion

The context in which a sound was presented to an awake primate systematically changed the filter approximating an auditory neuron. Neurons in primary auditory cortex of the awake primate increased in their selectivity and decreased in their responsivity as the sound input became noisier. Inhibition shaped the responses of A1 neurons only in noisy environments. Predictive experiments show that cortical representations of simple sounds require at least hundreds of neurons and that responses of neurons with inhibitory subfields are less predictable with linear techniques.

We thank K. O. Johnson, L. Miller, and K. Sen for useful commentary on the manuscript. R. Ramachandran and H. Attias contributed useful discussions about these issues.

This work supported by the Coleman fund, HRI, the Sooy fund, and National Institute of Neurological Disorders and Stroke Grants 1F32NS-10154 and NS-10414.

REFERENCES


J Neurophysiol • VOL 88 • DECEMBER 2002 • WWW.JN.ORG